ILI Benchmarking – Ensuring an Adequate Understanding of ILI Capability Under Real-World Operating Conditions.

Authors:

Aidan Charlton, Technical Manager (THEIA Digital Integrity & Innovation), Penspen, UK

Mohammod Rahman, Engineer (THEIA Digital Services), Penspen, UK

Abstract

Modern inline inspection tools offer an unequalled method to rapidly inspect pipeline systems for a range of damage mechanisms and have undoubtedly assisted reduction of incident rates within the pipeline industry. Cost of inline inspection can be high; in costs of preparation such as cleaning pig runs, increase of headcount on or in controlled installations, cost of reduced production in addition to the actual cost of the inspection by the ILI tool.

Inspection tool efficacy is therefore a key consideration, particularly in cases where secondary costs in addition to the actual ILI tool vendor cost are comparatively high. In these cases, the overall cost benefit analysis of ILI tool can give an increased incentive to contract the best technology and tool available. Commonly the data driving the cost benefit analysis is provided by the specific ILI tool vendor; so how can operators verify the claims of the ILI vendors and prove comparative ILI tool efficacy between vendors under real-world operating conditions?

Recent advances in computational capability of commonly available computer hardware has enabled increased capability in analysis conducted on large datasets. This includes the capability to conduct statistical analysis of every feature detected by multiple inline inspections across many individual pipelines.

The authors will detail a typical anonymised output of statistical analysis of multiple ILI vendors and ILI tools and will explain the insights that this can provide, irrespective of ILI vendor claimed specifications. The authors will demonstrate that the methodology allows an unbiased analysis of ILI tool capability under real-world operating conditions, and how this can be used to better inform the cost benefit analysis process used when selecting a specific ILI vendor and ILI tool.

1. INTRODUCTION

Penspen are a Pipeline Engineering Consultancy independent from all pipeline inspection vendors, therefore Penspen is fully agnostic to data-type and data-format.

Historical pipeline tools have progressed greatly from simple pipeline cleaning & proving tools into highly sophisticated Intelligent Inline Inspection (ILI) tools which can quickly traverse and inspect large distances within a pipeline system at high velocity and detect anomalies to a remarkable degree of accuracy.

Different vendors and technologies are available, all of which continue to innovate and improve; however, a pipeline operator must always decide which inspection vendor and tool offer the best combination of characteristics to detect credible threats to pipeline integrity. Is this the most accurate, the most reliable or even would a lower cost alternate tool prove acceptable?

Each operator will have different drivers and motivation when selecting a vendor, and a specific tool with which to conduct a particular inspection campaign. As the volume, quality and richness of much of the data being produced by inspection improves it is becoming increasingly difficult to perform objective analysis of results without improved data processing capability, so is the operator getting the best value from their selection?

To be able to automate the comparison and analysis of vast quantities of data from disparate sources and in differing formats we must harness developments in computing hardware such as GPU processing, and new computational methodologies to ensure that we make best use of the data available, and to make the most appropriate choices on which vendor is selected to generate the data in the first place.

As stated Penspen is agnostic as to data-type or data-format, and remains independent of ILI vendors. The objective of this paper therefore is to introduce a methodology which can be used to objectively and numerically compare real world reported results of many ILI inspections to give an indication of inspection efficacy based on comparison of features reported by various ILI vendors.

Many of the techniques detailed in this paper have been specifically developed by Penspen to allow next generation pipeline integrity assessments to be delivered directly to clients as part of the THEIA software package and build upon previous papers detailing fast automated alignment of ILI datasets & feature correlation.

2. OBJECTIVES

The overall technical objective is best described with a question: how can the accuracy of reported features from a particular ILI inspection be verified?

Signals data from ILI is not commonly transferred to operators, or acted upon by Engineers outside of the ILI vendor. The relevant data of interest for Engineers is the listing of reported features, what features have been detected and the dimension and surface location of features; in reality this information is an interpretation of raw signal data which is subject to variation of interpretation.

In order to resolve the technical objective, there are three distinct areas which must be investigated;

- Reporting the existence of features within the correct statistical probabilities,
- Reporting of feature dimensions within acceptable bounds,
- Reporting of surface location with good accuracy.

The reporting of surface location is evident where good feature correlation is obtained, and differences in reported feature location are compared using the statistical probability of a correct majority vote between all ILI datasets.

This is not discussed further other than to state on rare occasion a particular ILI does exhibit a high fraction of mal-reported feature locations when correlated features are compared across several ILI datasets, so is worth including for consideration of efficacy.

3. METHODOLOGY INTRODUCTION

Values to be reported by ILI are typically given a confidence level and a dimension or threshold, so for example a threshold for feature detection may be 80% of a specific feature aspect type will be reported at a nominal dimension of *x,y,z* dimension units. Features with differing aspect ratios will have varying degree of dimension dependant on the technology used. For example, a Traverse Flux Leakage tool can be expected to detect and size axially long features with better confidence and reduced tolerance than an Axial Flux Leakage tool, with the opposite being true for axially short features, both with respect to circumferential dimension.

The methodology must therefore take into account feature aspect ratio, and the specific claimed tolerances for detection and sizing of varying aspect ratio of feature. This results in a normalised dataset for further processing.

The first aim is to quantify the efficacy of reporting of features. To prove or disprove the efficacy of detection & reporting of features for a particular ILI comparison is made between correlated ILI datasets.

Figure 1 shows the theoretical comparison of a single feature detected by all of three ILI datasets. The absence of detection is equally valuable where the expected dimensions of the feature can be calculated for the temporal location missing. This is then used to backfill the dataset for further processing which can help establish if the non-reported feature has been plausibly omitted from the dataset within the stated reporting probabilities. Similarly, calculation for dimensioning efficacy of features can be conducted on normalised data.

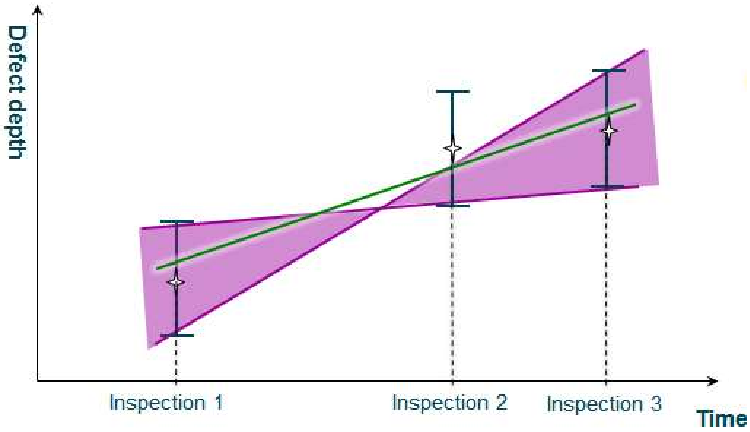

Figure 1: Example of comparison of multiple ILI for a single feature

It is important to note that a high degree of variability can be reasonably expected when inspecting the results of individual features, results must therefore be aggregated, and include sufficient number and quality of data points to give statistical significance to results. The methodologies are described in more detail in the following sections.

3.1 Methodology Phase 1: Feature reporting.

Naturally expected variation exists in over / under achievement of the actual feature dimension within a set of features. This applies to the reported variables: Depth, Width and Length. The plausibility of reported values can be mathematically modelled using typical regression models that account for variation in independent variables. Where two reported feature datasets with sufficient datapoint volume and quality are compared, an estimation of comparison of efficacy can be made.

Variation in feature dimension, such as caused by corrosion growth of metal loss features, for large datasets can be calculated using data from multiple inspections. Various previous papers describe this type of methodology, with the MDRLP methodology typically used internally by Penspen. Results are calculated using linear modelling and normal distribution for confidence levels, from this we are able to find the CGR. (P*Q – Q*gamma) / (alpha*beta – gamma^2)

Figure 2 indicates the initial computed CGR for a particular pipeline and Figure 3 indicates deviation from fit, in this case the mean truncated values have been utilised. Where differing technologies are to be compared the non-truncated values may be more appropriate, with the criteria being confidence in differentiation between reported status as either active corrosion or non-active mill-features. In cases where a technology has been used which cannot discriminate, all metal loss features are considered for all inspections in the comparison set.

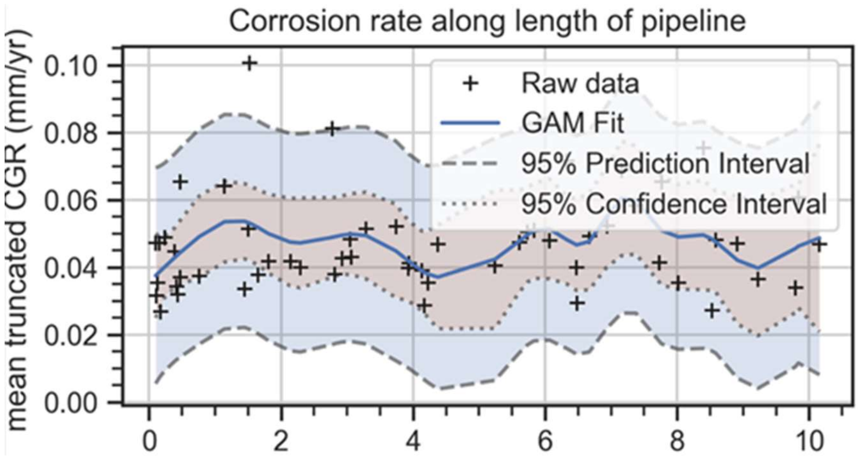

Figure 2: Example GAM fitted Mean Truncated variable CGR for whole pipeline.

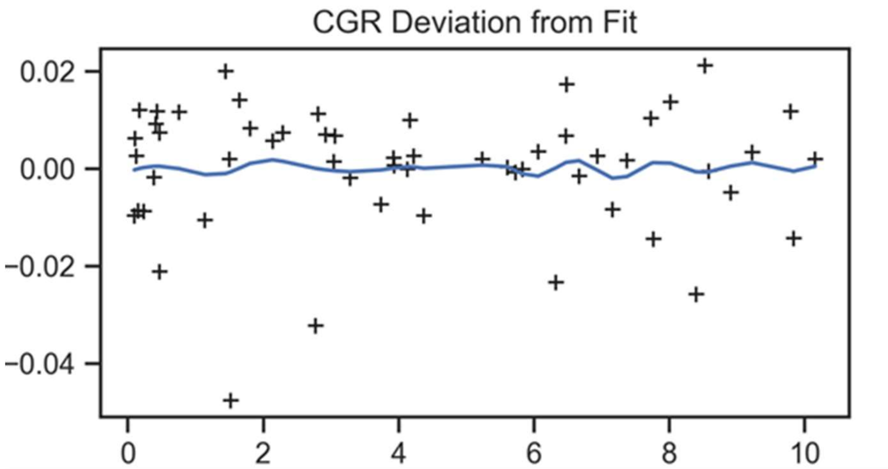

Figure 3: Deviation from fit.

A linear generalized additive model with error distribution is implemented to further factor in statistical variation of input variables. Generalized Additive Models (GAM) provide smooth semi-parametric models which combine additive models (non-parametric regression method) and generalized linear models (linear regression).
This is then utilised to generate a fit interpolation report using the GAM model and the computed CGR.

In instances where a particular ILI does not report a feature, it is possible to predict the missing dimension values with appropriate confidence and to justify the binary outcome; detected vs not detected. The aggregated results provide statistical analysis of the provided and imputed data values and comparison with the feature dimension at the appropriate probability of detection indicated by ILI tool tolerances & thresholds.

As a back-check for estimation accuracy of imputed values, the dataset of reported & correlated features is verified directly by systematically comparing reported values with estimated values for those instances where values are provided in all datasets. An example of this is indicated in Figure 4.

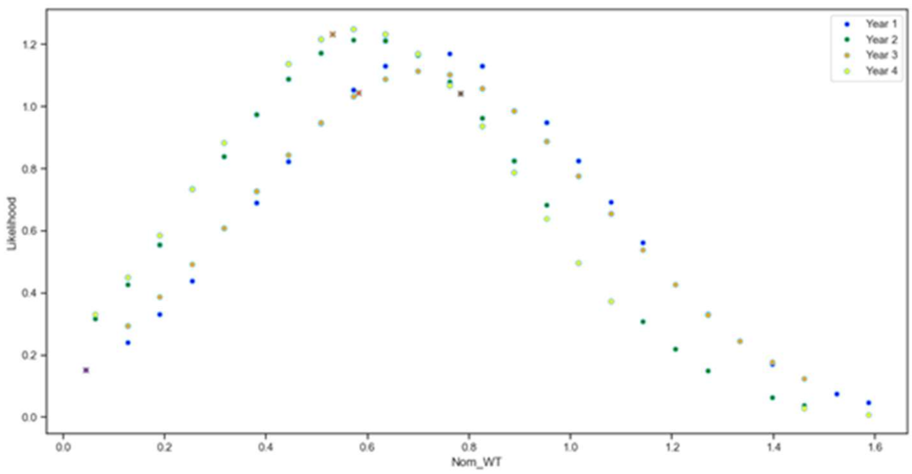

Figure 4: Example probability distributions for feature depths.

Table 1 and Table 2 illustrate an example dataset where four ILI datasets are considered, and appropriate null values inserted for back-testing purposes against imputed values.

	GW	A	B	C	D
0	230	NaN	0.1905	0.3810	0.6350
1	230	0.6350	NaN	0.7620	0.4445
2	230	0.4445	0.4445	NaN	0.2540
3	230	0.9525	0.8890	0.4445	NaN
4	230	0.6985	0.7620	0.1270	0.8890

Table 1: Example dataset with appropriate values set to null.

```
[[0.16468548 0.1905     0.381      0.635     ]
 [0.635      0.73676149 0.762      0.4445    ]
 [0.4445     0.4445     0.5100665  0.254     ]
 [0.9525     0.889      0.4445     0.48474645]]
```

Table 2: Example dataset with imputed values.

3.2 Methodology Phase 2: Feature Dimension Analysis

The ordinary least squares method is a linear least squares method for estimating unknown parameters in a linear regression model. The r2 score, correlation and RMSE of the model is utilized to further quantify whether a given inspection has under or overachieved against indicated thresholds.

All possible combinations are assessed using the ordinary least squares and statistical techniques/metrics to view the relation between the calculated/calibrated parameter values and measured values. By assessing all possible combinations as a parametric type study it is possible to aggregate the data to quantify the overall inspection.

Figure 5 indicates the comparison of linear regression fit lines where the depth of features is compared between inspection 1 and inspection 2. In this case a variation is evident between fit lines, and in this case the variation indicates that;

- Deeper features are reported to be deeper than expected by inspection 2.
- Shallower features are reported to be shallower than expected by inspection 2.

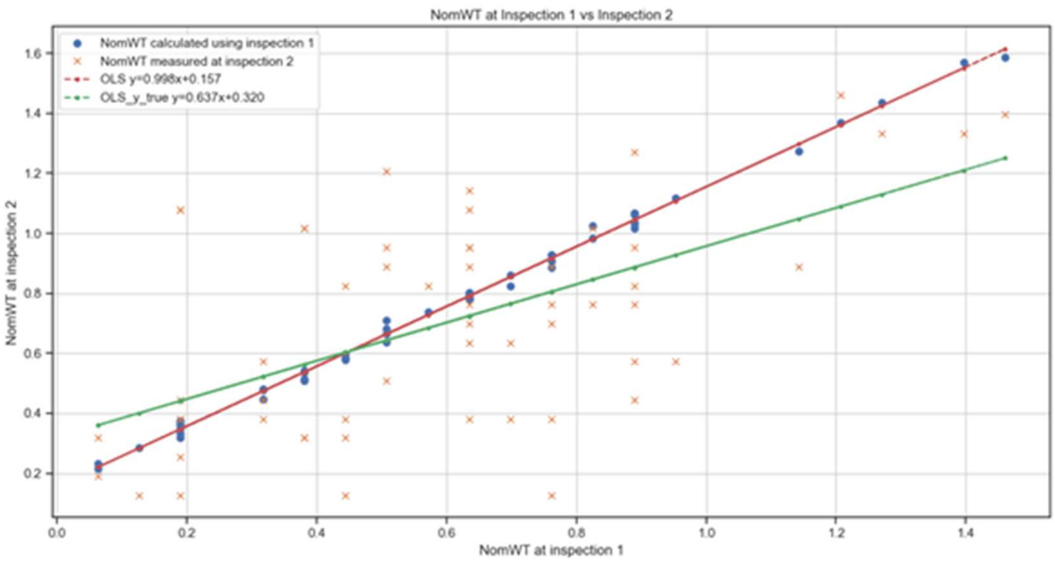

Figure 5: Predicted inspection 2 depth values vs measured Inspection 2 depth values.

From Figure 5 it is visibly clear that the RMSE indicates a poor correlation, and that the reporting of feature depth by inspection 2 would be of a lower efficacy than preferable, and that this is particularly evident in shallower features.

For comparison Figure 6 provides a similar output with relation to inspection 3 with comparison of expected and reported depth values. A higher degree of correlation is evident which results in an improved RMSE value giving higher confidence in accuracy of predictions.

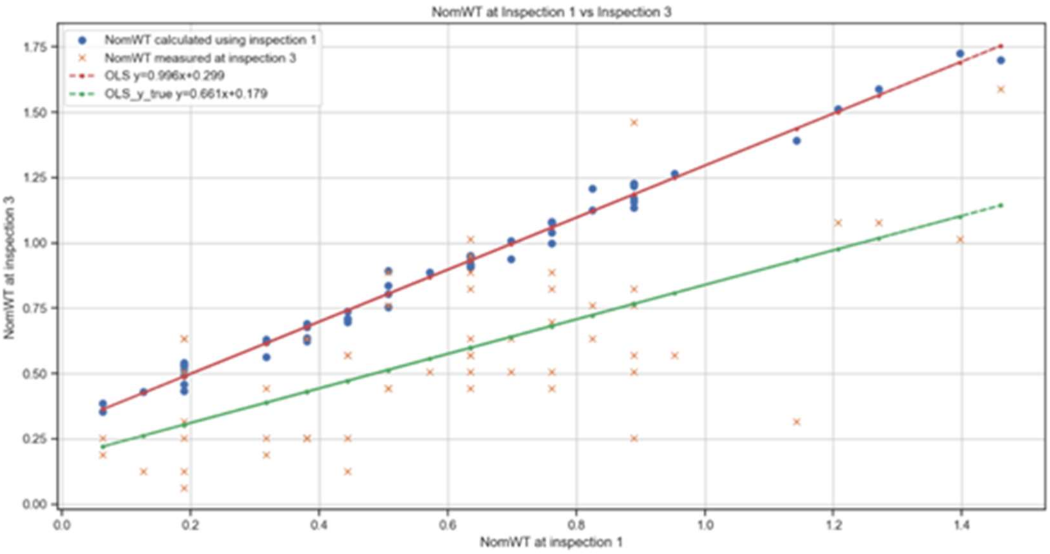

Figure 6: Predicted inspection 3 depth values vs measured Inspection 3 depth values.

We perform the regression analysis methodology for each possible dataset combination of variable dimension such that the efficacy of sizing of feature; depth, width & length is considered. The r2 score compares the fit of the model with the null hypothesis. The accuracy of ILI data can be found by calculating the change in different inspections combinations.

4. CONCLUSIONS

The best inspection tool for any particular pipeline or inspection campaign will vary dependent upon the particular pipeline and campaign specific factors.

For example; a launch from an offshore installation may have a higher emphasis placed upon a higher probability of successful completion on first inspection run, whereas a short onshore feeder pipeline may have a higher emphasis on efficacy of reported dimensioning.

This paper aims to describe a methodology which can be used to compare efficacy of reporting of features between various inspections and does not consider additional inspection priorities. This results in direct comparison of inspections under real-world operational conditions, no pull-test or other data collected under non-operational data is required.

The first part of enabling comparison of multiple ILI datasets is accurate axial alignment and feature correlation across all combinations of datasets. In the case of few datasets this is achievable manually. For large numbers of datasets to compare this rapidly becomes non practical given that the equation to assign combinations is a factorial based equation and expands the number of comparisons non-linearly.

Overall, more stable results are obtained by direct comparison of a greater number of datasets so ideally an automated axial alignment and feature correlation capability would be in place. Similarly, care must be taken not to infer the efficacy of particular technologies or vendors based on the comparison of datasets from few physical pipelines.

Ultimately it is possible to make statistical comparison of efficacy of reporting for various feature aspect ratios by vendor, tool or technology, for reporting each of, feature dimension, location and detection threshold; again this requires aggregation of large numbers of compared datasets to enable a suitable degree of confidence and to exclude the possibility of other factors which would otherwise be removed by large-scale aggregation.